DOCUMENT RESUME

ED 193 235 TM 800 407

AUTHOR Mitchell, Virginia P.; Smith, Richard M.
TITLE Determining Cut-Off Scores Using Rasch Ability Estimates.
PUB DATE Mar 80
NOTE 15p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Boston, MA, April 8-10, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Algebra; Basic Skills; College Freshmen; *College Mathematics; *Cutting Scores; Difficulty Level; Goodness of Fit; Grades (Scholastic); *Latent Trait Theory; State Programs; *Student Placement; Testing Programs; Test Items; *Test Validity; Two Year Colleges
IDENTIFIERS *New Jersey College Basic Skills Placement Test; *Rasch Model

ABSTRACT

The validity of cut-off scores, used to place students into English and mathematics courses, and a dichotomized success variable, representing student abilities related to success in three sequential mathematics courses, were investigated. This study, examining the cutting scores of students entering Mercer County Community College in Fall, 1978, on the state-mandated New Jersey College Basic Skills Placement Test, identified the success variable in a three course algebra sequence as an earned grade of C or above. Because the success variable for each course yielded the minimum ability required for success in that course, the raw score equivalent in each case was interpreted as a cut-off score. The Rasch model was used to link student success in a course with a particular ability on the latent trait. Items were calibrated using the BICAL program and adjusted. Results suggested that the abilities required for success in the three courses were extremely close to one another. Content and grading practices of the three course sequence should thus be redesigned. (BL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193235

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

DETERMINING CUT-OFF SCORES USING RASCH ABILITY ESTIMATES

VIRGINIA P. MITCHELL
SYRACUSE UNIVERSITY

RICHARD M. SMITH
MERCER COUNTY COMMUNITY COLLEGE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
V. Mitchell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

March 1980

Paper presented at the annual meeting of the National Council on Measurement in Education

Boston, April 1980

TM 800 407

ABSTRACT

In the Fall of 1978, Mercer County Community College, along with all two and four-year public institutions in New Jersey, began to administer the New Jersey College Basic Skills Placement Test (BSPT). In the process of changing from the previous placement instrument to the BSPT, the authors decided to investigate the validity of the cut-off scores used to place students into English and mathematics courses. In this study, a dichotomized success variable was included with a Rasch model-based calibration of several BSPT test items to determine what student abilities were related to success in three sequential math courses.

Introduction

The faculty and staff of Mercer County Community College in Trenton, New Jersey, have been concerned for several years about the difficulty of making correct decisions with regard to the placement of students into English and mathematics courses. Placement decisions have been more crucial lately with students' increased need for remediation in basic skills (College Entrance Examination Board, 1975) and more difficult with the increased heterogeneity of student population (Cross, 1971).

With the introduction of the New Jersey College Basic Skills Placement Test (BSPT) mandated by the Department of Higher Education to be administered to all students entering the state's public post-secondary institutions beginning in Fall, 1978, came the unique opportunity to overhaul the College's placement system and to analyze the assumptions underlying it. It was expected that such an investigation would have implications for the setting of BSPT cut-off scores for the placement of students into courses, as well as for the possible redesign of the content of several courses.

Background of the Study

For the past ten years, the placement of students into several English and mathematics courses had been done by means of the Comprehensive Guidance and Placement examination (CGP) published by the Educational Testing Service. Studies done during this period (Mitchell, 1976; Smith and Mitchell, 1979) have produced at best only moderate correlations between students' CGP at entry and course grades, suggesting that there are other factors besides English and mathematics ability, as measured by this test, which affect students' successful course completions. While this has also been shown to be the case with the BSPT (Smith and Mitchell, 1979), for the purposes of this study the variable

under investigation was mathematics ability as measured by the BSPT and a dichotomous "success" variable. Specifically, the relationship to be examined was that between students' mathematics ability as measured by the BSPT at the time of placement vs. that at the time of the completion of the course as indicated by the "success" variable.

This study represents a part of the analysis of the existing placement system centering on a three-course sequence in mathematics. These courses were "MA 109," or Elementary Algebra, "MA 122," or Intermediate Algebra, and "MA 103," College Algebra I. The first of these was required for most programs of study at the College and, consequently, was one of the most highly enrolled courses. It had been the focus of many redesign efforts including an attempt at adapting the content to a mastery model with required weekly tests, each of which the students were not allowed to take without passing all preceding ones. In spite of all this the low success and high attrition rates in MA 109 persisted and even worsened. The seriousness of the problem and the pressure on the students was increased by the decision made by several departments that an acceptable grade, i.e., "C" or better, was required in MA 109 before upper level course work in various programs could be pursued.

Thus, the most dramatic problems centered around the first course in this sequence. This is not meant to imply that students' success rates were acceptable in the other two, however. It seemed logical that, since students could be placed directly into any one of the three on the basis of BSPT scores, and that the course content for all three was in the area of algebra, that they should all be examined.

During a previous investigation of the utility of the Rasch model in the improvement of the predictive ability of the CGP (Smith and Mitchell, 1979), the items on the Computation, Applied Arithmetic, Elementary Algebra, and Intermediate Algebra Sub-tests were calibrated using the BICAL program (Wright and Mead, 1977) on the scores of students entering the College in Fall, 1977. This preliminary study suggested the potential utility of the Rasch model with standarized placement instruments. Most items were found to fit the model indicating unidimensionality of the mathematics ability trait, and, when the few misfitting items were excluded for the purpose of re-estimating abilities, the item difficulties were found to be stable. Thus, ability estimates were produced which were on the same scale with the same origin as the original estimates.

The Rasch model had been employed for several reasons. Since the nature of these sub-tests suggested that their purpose was to assess a basic underlying trait, i.e., mathematics ability, it seemed reasonable to explore a latent trait interpretation of the placement problem. In particular, the Rasch model offered the following features: item difficulties and ability estimates are calibrated in the same metric know as "logits"; item difficulties are "sample-freed" and person abilities are "item-freed" (Waller, 1976); and a linear scoring metric is employed (Wright, 1977). Also, it would be highly desirable to be able to predict the probability of a student with a certain estimated ability being able to answer an item with a given difficulty correctly. That is, because both abilities and difficulties have a "location" on the given latent trait and these locations are ordered and linear, a great deal more information is available to the test user. It is just this kind of information which is important in making placement decisions, i.e., whether or not a person with a certain ability is likely to know a particular content area.

Description of the Study

Developed over the course of a year by several committees composed of subject matter and testing specialists, the BSPT was intended to provide a common measure by which basic English and mathematics abilities of all students entering public post-secondary institutions in the state could be determined. While colleges and universities were not permitted to use scores on the BSPT to make admissions decisions, they were encouraged to use the test for placement purposes.

The calibration of the Computation and Elementary Algebra Sub-tests of the BSPT mentioned above made use of the scores of students entering the college in Fall, 1978, to produce an item bank for the placement of students into three sequential mathematics courses. Based upon a discussion of the Rasch model by Wright (1977) in which he suggested that such an item bank could be criterion referenced by the inclusion of a dichotomously scored success variable, it was decided that it would be useful to include such a variable in this bank. "Success" was therefore defined as an earned grade of "A," "B," or "C," while "failure" was considered to be a "D," "NC" (no credit), or a "W" (withdraw). This investigation, therefore, represents an attempt to determine how well the success variable would fit the model and how the three courses would compare with each other on this variable. Because the success variable for each course yields the minimum ability required for success in that course, the raw score equivalent in each case can also be interpreted as a cut-off score, possibly for placement purposes.

In the calibration process, items which are answered either correctly or incorrectly by everyone in the sample under consideration furnish no information for the estimation of abilities and are, therefore, excluded. Also excluded are individuals who either answer all or none of the items

correctly. In addition, BICAL permits the user to designate a chance level below which scores will be excluded from the calibration. Among the three calibrations only one individual was excluded from consideration, and this was for the MA 109 group because of a score below the designated chance level of 8.

Since only the Computation Sub-test was used for MA 109 placement, the thirty items from this test were dichotomized into "right-wrong" responses for those students who registered for the course in Fall, 1978. For MA 122 and MA 103, however, all sixty items from the Computation and the Elementary Algebra Sub-tests were dichotomized, as both tests were used for placement. In all cases, grades from students enrolled in these courses for Fall, 1978, semester were also dichotomized and were calibrated along with the items by means of the BICAL program. A fit statistic with an expected value of "1" was also generated for each item which, with its associated standard error, indicated how well each item including the success variable, fit the model.

Because the item calibration process placed the origin point at the mean of the item difficulties, the results of the three calibrations had to be adjusted by suitable constants in order that all items could be compared on a common scale. Two techniques are generally employed to determine the adjustment constants. The first is merely an inspection of the scatter plots of the item difficulties for two calibrations at a time. The second involves the computation of the mean of the differences in the item difficulties of corresponding items. Both techniques were used in order to compare them. Scatter plots of MA 109 and MA 122 item difficulties for the Computation Sub-test and of MA 122 and MA 103 item difficulties for the Computation and Elementary Algebra Sub-tests were produced, as were the means of the appropriate item difficulty differences.

Results/Conclusions

The MA 109 calibration group contained 122 subjects after one was excluded for having a score equal to the designated chance level of 8. There were 128 subjects in the MA 122 calibration sample and 72 in the MA 103 group. The question of how many subjects are required to produce stable item difficulties has been addressed by Wright (1977) who said that calibration samples of as few as 100 can be useful on tests of more than 20 or 30 items. The standard errors of the "success" items were found to be .21 logits for MA 109, .19 for MA 122, and .26 for MA 103 which are sufficiently close to the acceptable value of .25 logits cited by Wright to provide the needed stability.

Only 3 of the 30 items on the BSPT Computation Sub-test were found to have a total fit statistic of more than 2 standard errors (SE=.13 logits) from the expected value of 1 when calibrated with the MA 109 group, and only 1 of these was more than 3 standard errors away. In addition, 4 items were more than 3 standard errors from the expected value for the between-group fit statistic (SE=.58 logits) which is generated through BICAL for the analysis of the behavior of items with 6 user-specified subgroups which are sequential from low to high mean abilities. Two of these 4 items had acceptable total fit statistics, but a careful analysis of the test should include an examination of items such as these which fail to fit the model for certain subgroups of person abilities.

Finally, the success variable for MA 109 fit the model quite well with a total fit statistic of 1.05 logits and a between-group fit of .55. Its item difficulty of 1.33 before adjustment to a common scale was within 1 logit of the group's mean ability of .45 logits. This was found to correspond to a raw score on the Computation Sub-test of between 23 and 24 out of 30 items.

For the calibration of the Computation and Elementary Algebra Sub-tests on the MA 122 group, 7 of the 60 items had a total fit statistic of more than

2 standard errors from the expected value of 1 (SE=.13 logits), 6 of which were more than 3 standard errors away. Three items, 1 of which had an acceptable total fit statistic, had between-group fit statistics which were more than 3 standard errors from 1 (SE=.58 logits).

The success variable for this calibration had a total fit statistic of .9 logits and a between-group fit of .43 logits. The item difficulty was .58 logits as compared with the mean ability for this group of .09 logits. This was equivalent to a raw score of 37 out of 60 items.

The calibration for the Computation and Elementary Algebra Sub-tests on the MA 103 group produced 8 items with total fit statistics which were more than 2 standard errors from 1 (SE=.17 logits), only 4 of which were beyond 3 standard errors. Four items had between-group fit statistics which were more than 3 standard errors from 1 (SE=.58 logits), all of which also had total fit statistics which were 2 standard errors beyond 1. None of the items with extremely high between-group fit statistics were the same for both the MA 122 and MA 103 calibrations, and only one of the items which misfit by 3 or more standard errors on the total fit statistic were the same for these calibrations.

The success variable for MA 103 was again found to fit the model with a total fit statistic of 1.04 and a between-group fit of 1.41. The item difficulty of .82 logits was well within one logit of the group's mean ability of 1.10 logits. This corresponded roughly to a raw score of 40 out of 60 items.

When the item difficulties from the MA 109 calibration of the Computation Sub-test were plotted against those from the MA 122 calibration, an inspection of the points suggested an adjustment constant of about .8 logits. The computation of the mean difference across item difficulties for corresponding

item produced an extremely close result, .75 logits. Similarly, the inspection of the MA 122 - MA 103 scatter plots suggested that no adjustment of the scales was needed, and the computation produced a mean of less than .01 logits. Thus, by subtracting .75 logits from the item difficulties produced for MA 109, the items for all three calibrations could be interpreted by means of the same scale.

When the scales were aligned, it was found that the log abilities of the success variable were .58 for the MA 109 calibration, .58 for MA 122, and .82 for MA 103. Thus, the ability required for success in MA 109, Elementary Algebra, was virtually equivalent to that for MA 122, Intermediate Algebra, and only 1.14 standard errors from that required for MA 103, College Algebra I. These differences correspond to a raw score span of from 37 to 40 units on the BSPT.

This suggests that the abilities required for success in the three courses are extremely close to one another which in turn leads to many other questions. For instance, does success in the first-level course really require the same ability as that required for the second-level course, i.e., are they equally difficult or easy? This question should logically lead to an evaluation of the two courses including aspects such as content scope; content analysis of textbooks, study guides, and tests; and student evaluation requirements. Such an investigation could also be extended to the third course, MA 103. The main question really should address the nature of the material to be covered in the three courses. If it is truly hierachical in nature and can reasonably be divided among the three courses, careful planning should go into a redesign or re-examination of classroom presentations, material, and evaluation procedures so that beginning and intermediate students won't be unnecessarily frustrated in their efforts to master the content.

Importance of the Study

The determination of cut-off scores by means of examining the ability levels of students who succeed in courses with the abilities of those who fail is intuitively appealing and logical. The problem until now has been that estimates of abilities have always been dependent on the particular items the student was tested on, just as the difficulties of the items depended on the examinee sample. Since the Rasch model frees difficulty and ability estimates from these constraints, it now becomes possible to link student success in a course with a particular ability on the latent trait. When, as in this study, success does not clearly differentiate the three groups, this says that something very basic is wrong with the math course sequence. Either the assumptions about student abilities, especially for those placed into the beginning course, need to be re-examined or course content or grading procedures should be redesigned.

In addition, the use of the Rasch model allows us to estimate a student's probability of success on an item. When the items's difficulty matches the student's ability, the student's probability of success is 50%. The use of the "success" variable, then, indicates what a student's ability would have to be in order to have a chance of earning an "A," "B," or "C" in one of the three mathematics courses. This information could not be obtained by using raw scores alone.

TABLE 1

SUMMARY OF RESULTS

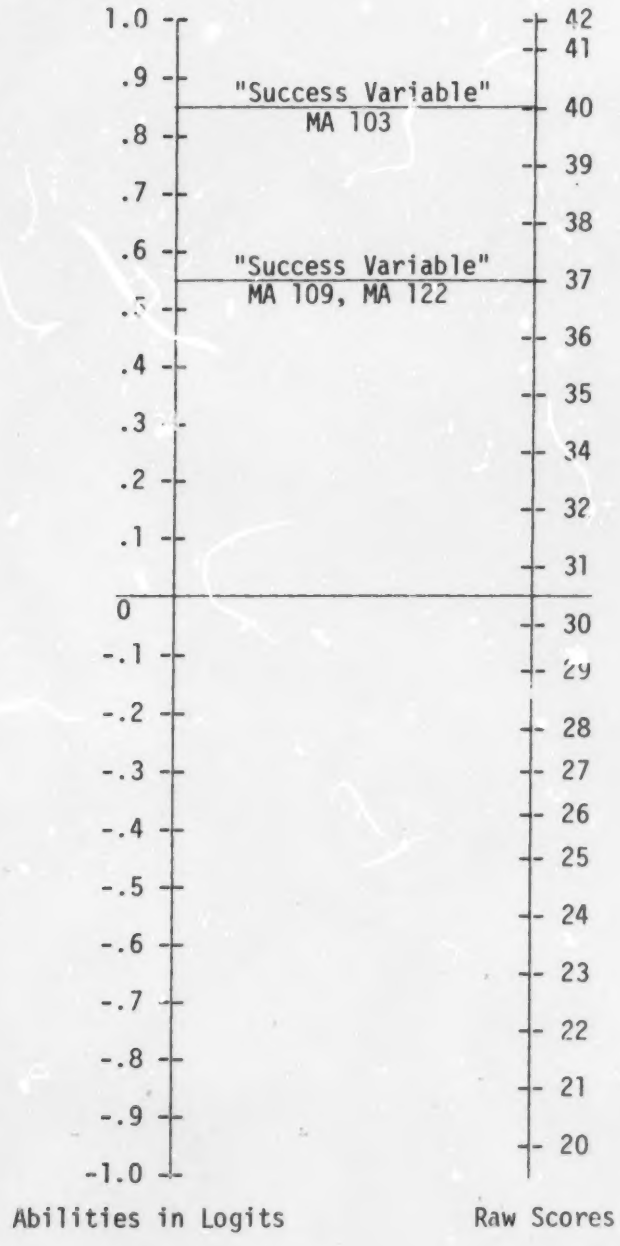
	MA 109	MA 122	MA 103
Number of subjects	122	128	72
BSPT Subtest(s) Calibrated	Computation	Computation Elementary Algebra	Computation Elementary Algebra
Number of Items	30	60	60
Number of Items More Than 3 SE's from Expected Value of Total Fit Statistic	1	6	4
Number of Items More Than 3 SE's From Expected Value of Between-group Fit Statistic	4	3	4
"Success" item:			
Standard Error	.21 logits	.19 logits	.26 logits
Total Fit[1]	1.05	.9	1.04
Between-group Fit[1]	.55	.43	1.41
Item Difficulty	1.33[2] logits (.58)	.58 logits	.82 logits
Raw Score Equivalent	23-24[2] (37)	37	40

[1] These are f-statistics

[2] Prior to adjustment to the common scale. Figures in parentheses are post-adjustment.

TABLE 2

PORTIONS OF THE COMMON LOG ABILITY AND RAW SCORE SCALES
FOR ALL THREE MATH COURSES



References

Cross, K. P., Beyond the Open Door: New Students To Higher Education. San Francisco: Jossey-Bass 1971.

College Entrance Examination Board, College-Bound Seniors, 1974-75. New York: College Entrance Examination Board, 1975.

Hambleton, R. K. and L. L. Cook, Latent Trait Models and Their Use in the Analysis of Educational Test Data. Journal of Educational Measurement, 1977, vol. 14, no. 2, 75-96.

Hambleton, R. K., H. Swaminathan, L. L. Cook, D. R. Eignor, and J. A. Gifford, Developments in Latent Trait Theory: Models, Technical Issues, and Applications. Review of Educational Research, 1978, vol. 48, no. 4, 467-510.

Mitchell, V., C. G. P. Validity Study: Entering Students, Fall, 1975. Office of Institutional Research, Mercer County Community College, 1976.

Rever, P. R. and L. K. Kojaku, Access, Attrition, Text Scores, and Grades of College Entrants and Persisters: 1965-1973. Research in Higher Education, vol. 5, no. 4, 1976, 289-299.

Smith, R. and V. P. Mitchell, Improving the Predictive Ability of Placement Tests Using the Rasch Model for Scoring. Paper presented to the Annual Conference of the National Council on Measurement in Education, San Francisco, 1979.

Waller, M., Ability removing random guessing model. Paper presented at the Annual Conference of the American Educational Research Association, San Francisco, 1976.

Wright, B. D., Solving Measurement Problems with the Rasch Model. Journal of Educational Measurement. 1977, vol. 14, no. 2, 97-116.

Wright, B. D. and R. J. Mead, BICAL: Calibrating items and scales with the Rasch Model. Research Memorandum Number 23A, University of Chicago, 1978.